REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From – To)*
20-02-2009	Final Report	12 September 2007 - 12-Sep-08

4. TITLE AND SUBTITLE

The Form and Function of Linguistic Elements

5a. CONTRACT NUMBER FA8655-07-1-3094

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Professor Emmon Bach

5d. PROJECT NUMBER

5d. TASK NUMBER

5e. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of London
22 Coniston Road
Muswell Hill
London N10 2BP
United Kingdom

8. PERFORMING ORGANIZATION REPORT NUMBER

N/A

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

EOARD
Unit 4515 BOX 14
APO AE 09421

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

Grant 07-3094

12. DISTRIBUTION/AVAILABILITY STATEMENT: Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This report results from a contract tasking University of London as follows: In the two papers that constitute the formal portion of the project, the DoubleR Model will be assessed in light of relevant proposals from other current formal theories. As the goal at this stage is to develop the mapping between form and function, the candidate theories/issues for initial consideration will be those that address the nature of the syntax-semantics interface, both in its constructional aspects and its systematic, compositional domains. Beyond the basic level, aspects of meaning relevant to information structural content (topic, focus, contrast and backgrounding) may become relevant, and may help explain some of the results found in Dr Ball's initial study.

Paper 1 (Formal) will consist of a general overview of current syntax-semantics research and issues relevant to DoubleR Model and dataset. This will be presented for discussion with the expectation that two or three topics or theories will be selected for more intensive investigation at stage 2. Some likely candidates for examination at this stage include formal approaches in syntax and semantics such as combinatory categorial grammars, 'cartographic' and 'flexible' grammars from the generative tradition, formal accounts of the interaction between syntactic and information structure. These all address issues of multiple categorial structures and the way in which different types/aspects of interpretation can be conveyed by a given utterance in a discourse context.

Paper 2 (Formal) will consist of an in-depth review and assessment of selected approaches and topics in the light of theoretical issues posed by the DoubleRTheory and other construction grammar approaches. The end deliverable will be a formally written paper contrasting, comparing and critiquing these theories from the formal, discourse-informational and ognitive perspectives. As a result of this analysis and critique, it is expected that Dr. Ball will be able to further fine tune his theory, resulting in the best possible model to implement in the VERBOSE Project.

15. SUBJECT TERMS

EOARD, Psychology, Cognition, Linguistics

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18, NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
UNCLAS	UNCLAS	UNCLAS	SAR	50	TAMMY SAVOIE, Lt Col, USAF
					19b. TELEPHONE NUMBER *(Include area code)* +44 (0)1895 616459

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39-18

Formal Systems for Representing Changing Situations

Emmon Bach, PhD, and Wynn Chao, PhD

in partial fulfillment of EOARD Grant # 073094

Form and Function of Linguistic Elements

Principal Investigator: Emmon Bach

Final Report 20080215 for EOARD project

Dynamic Information Systems: Notes on some systems of grammar and interpretation.

.

Summary:

We deal with several different setups for specifying the syntax/grammar and interpretation for natural languages, together with notes on implementation and interfacing with online processes for dialogue, etc. Our descriptions are drawn in broad strokes. We give names for systems that are suggestive of actual frameworks and theories currently in use, but without many formal details.

Contents:

2. Construction Grammar

3. Miminalist Program Grammars

4. Categorial Grammars

III. MetaTheory

IV. Information systems

V A sample simple set up:

VI. Information states and systems of grammar and interpretation.

VII. Kinds of exchanges

VIII. Some samples

IX. Some Options

X. Where to go from here

References

Formal Systems for Representing Changing Situations

Emmon Bach, PhD, and Wynn Chao, PhD

in partial fulfillment of EOARD Grant # 073094

Form and Function of Linguistic Elements

Principal Investigator: Emmon Bach

Final Report 20080215 for EOARD project

Dynamic Information Systems: Notes on some systems of grammar and interpretation.

.

I Overview

We deal with several different setups for specifying the syntax/grammar and interpretation for natural languages, together with notes on implementation and interfacing with online processes for dialogue, etc. Our descriptions are drawn in broad strokes. We give names for systems that are suggestive of actual frameworks and theories on the market, but without many formal details.

These notes are pointed toward two tasks:

(1) processing linguistic inputs to yield semantic interpretations;

(2) using these interpretations plus contextual and other resources to produce data representations in the form of Information States and structures built from them, such as dialogues, logs, and the like.

We will refer to the class of systems that we have in mind as **Dynamic Information Systems (DIS)**.

It is assumed that these two tasks are necessary steps in developing systems of information exchange among (human and robotic) agents.

First, we list some very general assumptions and dimensions of choice.

II. General Assumptions and Options.

A. Components that are necessary for any setup.

1. grammar/syntax

We mean here the part of a system that is invoked online as opposed to the lexical and encyclopedic knowledge that goes into making and analyzing lexical items

2. phonological or graphical component.

This is the interface part of a system, based on the medium that is used to record and transfer information.

The choice depends on whether there is to be an interface to audio input/output or written input/output. This is the part of a linguistic system that makes a physical interface between users. We will couch our discussion in terms of written input and output, understanding that a live audio system would have to overcome major input-output problems of speech recognition and production.

3. Lexicon

This is the repository of the basic elements that are put together or recovered in the synthesis or analysis of linguistic materials.

4. Meaning components:

model-theoretic semantics / dynamic semantics

pragmatics: context theory

implicatures: conventional / conversational

presupposition

etc.

The use of "etc" here and throughout this report is meant to signal a principled choice for systems that are somewhat open-ended, so that distinctions and material that might not fit into a neat set of pigeon-holes is not lost but remains "there" for possible future incorporation.

5. Processor(s) for texts (dialogues etc).

Here we mean the parts of a system that deal directly with input and output, including parsers or production tools, lookup routines for accessing the lexicon and other repositories of information.

6. Comment

Different frameworks embody different -- sometimes sharply different -- views of how these linguistic subsystems are articulated and related. Many frameworks maximize the domain of the lexicon and some deny that there is any principled basis for distinguishing between the lexicon and the rest of the descriptive apparatus.

B . A sampling of frameworks:

Some of the options for grammatical descriptions on the marker currently are those associated with these headings:

principles and parameters models

construction grammars

categorial grammars

miminalist program grammars

brief mention of some other setups: Dynamic Syntax, LFG, HPSG

Comment: Some researchers claim that there is no need for a grammar, or that the grammar is an artifact or reflex of a parser, or derivable directly from a parser, or other kind of processor. We need to cast a net wide enough that such choices can be seen as particular instantiations of a general scheme.

We describe briefly and informally several of the frameworks just mentioned.

1. Principles and Parameters Models.

The Principles and Parameters framework characterized a large amount of work in syntax in the 80’s and 90’s of the last century in the Chomskyan line of work. It was meant to break from the earliest work in the generative-transformational line that began with Chomsky (1957) and continued into the frameworks from around Chomsky's Aspects (1965). The main thrust was to move from grammars centered on covering the details of particular languages toward more nearly universally applicable systems that relied on stating general principles (constraints on rules, etc.) and particular "parameters" that could be set in one or another way to achieve the observed variations among languages. Some signal general works in this kind of approach were Chomsky's books and papers of the time (for a critical review of this period of what the authors call Mainstream Generative Grammar, see Culicover and Jackendoff 2005). A specification of a model-theoretic semantics linked to such systems can be found in Heim and Kratzer (1998).

Two influential streams of research from roughly this period are those stemming from Richard Kayne and Guglielmo Cinque. A valuable collection of work in both these lines is Cinque and Kayne (2005). Cinque's work and work inspired by it is especially relevant in the current context, as it projected a universal set of functional categories and their relationships including many that are directly connected to categories of discourse, situation, time and place contexts, and the like.

Early generative grammar discussed traditional constructions like Passive, Raising to Object / Subject and so on. In later developments of the Chomskyan line, such units of analysis were abandoned as artifacts or epiphenomena and their features reconstructed as the effects of smaller operations and attributed to the effects of constraints on general operations such as "move-α" that were assumed to apply freely whenever the appropriate configurations for their application were met. This kind of move was countered early on in the framework and theory of construction grammar (Fillmore et al. 1988, Goldberg 1995).

2. Construction Grammar

The construction grammar program puts forward one main idea: the units of analysis are not organized into a single hierarchy but rather can be thought of as “constructions” of varying scope and generality, sometimes tied to individual words and patterns of words and other linguistic elements. So the early basic paper just mentioned (Fillmore et al., 1988) was devoted to the “let-alone” construction which can only be interpreted within a certain class of contexts. A language is to be characterized as a collection of constructions, ranging in generality from the kind that correspond to the rules or constraints that say that (many) sentences consist of a subject and a predicate to ones like the “let-alone” patterns or the “way” constructions discussed by Goldberg.

The construction grammar program has put forward a number of claims or planks, many shared by cognitive grammar. One common theme is that complex phenomena cannot be neatly compartmentalized according to the traditional rubrics of syntax, semantics, lexicon, grammar, and so on. The opposite strategy is to say that complex phenomena are best accounted for as an interaction among separate subsystems, each dealing with a narrower range of principles and effects. As an example, take interactions between truth-conditional or model-theoretic semantics and pragmatics as (at least) context theory. It is not clear that lumping these two domains into a single theory is better than trying to account for them by separate (but connectable) subsystems.

In spite of a lot of propaganda against formalization, any use of a system for a computationally accessible purpose requires precise specification of some sort. Implementation or formalization of construction grammars has leaned toward some system such as HPSG. Below we will attempt to sketch a specification of a flexible system based on a different tradition, that of categorial grammar.

3. Miminalist Program Grammars

Work by Chomsky, starting around 1995 (Chomsky 1995) represents a move toward radically simplified systems. A number of writers have worked toward formalizations that bring out the strong resemblances to categorial grammars (Lecomte 2005, Lecomte and Retoré 2001, Stabler 1997). In some ways, the systems suggested by minimalism come very close to the systems that have come from the tradition of Montague Grammar. We will not go into the Miminalist Program here.

.

4. Categorial Grammars

Categorial Grammars (CG's) were the earliest generative grammars in the modern era, with the first explicit discussion of them in the work of Ajdukiewicz (1935). They were taken up by Bar-Hillel, Lambek, and Curry in the sixties, and there has been a fair burgeoning of the tradition in the last two decades and a half (Oehrle et al. 1988 is a good source for work in the initial part of this revival of interest). Richard Montague used ideas from categorial grammar. Of importance for us here is that the systems are straightforwardly related to parsers, as they lend themselves exceptionally well to incremental ("left-to-right") interpretation. We lean toward the developments known as Combinatory Categorial Grammar (CCG: Steedman 2001, 2005). There is a considerable literature on parsing using such frameworks.

Although we will not try to spell out an implementation here, when we discuss below systems of Information States and dialogues, we will think about the crucial step of mapping natural language utterances or scripts into semantic representation as carried out within some such framework.

III. MetaTheory.

To compare theories or frameworks it is necessary to find a general enough set of terms and ideas that particular theories can be formulated as choices within such a general scheme or **metatheory**.

We take Richard Montague's Universal Grammar supplemented with a pragmatic theory as such a general framework. This sense of Universal Grammar is completely different from the Universal Grammar of the Chomskyan tradition. The former is something like a set of options for setting out the form and content of any theory that is about languages in the most general and abstract sense, while Chomsky's UG is intended to characterize whatever it is that humans have as a capacity or potential for learning and using natural, that is, human languages. (This view is close to the view of Terrence Deacon's "semiotic constraints" (Deacon, 2003), compare also Hockett's definitional universals (Hockett, 1963/1966.)

We lay out now the notion of a formal grammar (synonyms: recursive grammar, generative grammar).

Components: lexicon, rules, linking principles.

A formal grammar G specifies/generates/defines a language L(G): a set of signs.

Each sign is a k-tuple of objects, including, for example, these item:

syntactic representation

semantic interpretation

phonological reresentation

phonetic interpretation

(graphemic representation)

Further layers are possible: for example a semantic interpretation might itself be a multiplex object including implicatures of various kinds, contextually definable items etc.

The view laid out here is quite traditional, compare Saussure's notion of a sign, and the view of a grammar as a system linking sounds (or more generally forms) with meaning. The basic view of linguistic items as collections or sequences of multiple objects is common to many current frameworks such as HPSG, various Categorial Grammar theories. A more recent instantiation: Chomsky's minimalist program (Chomsky 1995 et seq.).

Particular instantiations of this general framework arise by making specific choices, for example:

What are the properties of the syntactic representations?

labeled trees / phrase-markers

bare strings of symbols

etc

What are the properties of the semantic interpretations?

model structure: individuals, truth values, worlds, times, situations

Static vs Dynamic systems (see below)

Types of linking: there are two main choices:

configurational

rule-to-rule

The **configurational** view has been the favored choice in most explicit (generative) grammars in the tradition of Chomsky. The grammar is thought of in the first instance as enumerating sets of representations: classically, phrase markers in the usual sense of the term: labeled (proper) trees, “logical forms” (LF) as representations of the predicate argument and quantificational structure of the meaning of sentences and other syntactic objects. In standard generative grammars of the Chomskyan tradition, sets of phrase-markers were enumerated at various “levels” and related to each other by mappings from one sort of object to another: Deep Structures, Surface Structures, Logical Forms (and on into phonological representations), later Semantic or Conceptual Structures (Jackendoff 1990, and much subsequent work).

This mode is to be contrasted with the **rule-to-rule** mode associated especially with work in the Montague tradition. Here a grammar is thought of as a recursion starting with a

lexical base and proceeding by means of constructive **rules** to build complex expressions, each rule specifying the input categories, the resultant category, and operation of the appropriate sort to yield the output, and -- crucially here -- the **interpretation** of the resultant expression as a function of the interpretations of the input categories. This is what is meant by the designation **rule-to-rule** (Bach 1976).

Example: (modeled on PTQ): Subject-Predicate Rule:

> If α is a member of the set of TermPhrases and β is a member of set of IntransitiveVerbPhrases, then γ is a member of the set of Sentences, where γ is the concatenation of the NOMINATIVE of α with the AGREEMENT-with-α of β. The interpretation of γ is the value of the interpretation of α applied to the interpretation of β.

This example can be taken to show how a parallel constructive grammar (as we may call it) works. We ignore the complication of intensional meanings that are actually used in Montague's work. We mainly want to show how this rule-to-rule approach differs from the configurational set-up of many semantic theories associated with phrase-structure theories. There the interpretation is based on some analysis into trees or the like. One consequence of taking the parallel approach is that the interpretation is completely independent of the particular operations used by the rule. In this respect, the setup is reminiscent of Lexical Functional Grammar, where phrase structure rules are associated with the construction of Functional Structures which then are the basis for semantic rules.

Example:

A clear instance of the difference between configurational and parallel approaches comes with the socalled bracketing paradoxes of examples like *set-theoretical*. On the one hand, the morphological structure of the word seems to go like this:

set-theoretical : set +[[theory + etic] –al]

But semantically the word seems to rather come from putting the compound *set-theory* together with an adjective-forming item *–[etic]al*. So from a configurational point of view there is a mismatch between the semantic structure and the morphological structure. But from the parallel rule-to-rule point of view there is no reason not to derive the adjective by applying a rule to the compound *set-theory* to derive the adjective in a way that respects the semantic structure. Obviously, then there is a trade-off between the system of allowable construction rules and the semantic options (a point made by Dowty in a recent publication, 2007).

Classical Montague grammar is an example of a **static** system: intensional functions are functions from world/time pairs and interpretations are given relative to assignments of values to variable, so that there is a degree of dynamism already in the system. More is added if the contextual elements are extended to include, for example, speaker, place, and so on.

A more thorough-going recasting of model-theoretic systems comes with the **dynamic** theories of Kamp, Heim and others (Heim 1982, 1983, Kamp 1981, Kamp and Reyle 1993; see Chierchia, 1995 and further work referred to there). Here, in the first instance, linguistic objects are thought of as functions from contexts to contexts (e.g. sets of assignments of values to variable). Classical semantics is then recovered in conditions of embeddability of DRS's (Discourse Representation Structures) into a model, and so on.

Examples of the naturalness and utility of such approaches are easy to come by in natural languages. Consider a sentence like (1):

1. Every student passed the examination.

Classically, this sentence would count as true if and only if for every x: x a student, x passed the examination. Leaving aside the proper interpretation of the "the examination" and the past tense, and providing a bit of context, we normally understand the sentence to refer only to students in the relevant situation:

2. The physics class had an examination yesterday. Every student passed the examination.

This example provides another case to be solved: the apparent binding of the value of "the examination" across sentences, parallel to the problem of anaphora across sentences:

3. A student entered the room. She walked to her seat.
4. ??Every student entered the room. She walked to her seat.

Ordinary language texts, and dialogues in particular, are shot through and through with such phenomena. We conclude that something like dynamic semantics must be part of any reasonable candidate information system.

As a start, we may assume that the model structure includes a set of situations. These may be thought of as a separate set or as a replacement of the set of possible worlds, with classical worlds taken to be maximal situations (Bach 1981, 1986; Kratzer 1989, 2007).

Accordingly, the set A of individuals in the model is refined into sets A(S), A(S')...where A(S) is the set of individuals in the situation S. These sets of individuals are conveniently thought of as divided into two families: those that are given by the initial specification of a "common ground" (Stalnaker, 1978) and those that are part of the locally developing context. So for example the individual denoted by the phrase "the Pythagorean theorem" is part of the common ground. Similarly, for other elements of the denotation: "students" refers to the set of things that are students, while "the students" picks out a set of elements in the situation domain that are students (and salient, etc.).

Further elaborations of the model structures include all sorts of Sorts, as in work of G. Carlson (Kinds, Stages, Objects), G. Link (Mass and Plurals), additional sets such as Properties in Chierchia's work.

Some properties of the grammar.

As noted above, we assume a basic split between Lexicon and Grammar (proper).

The Lexicon is the base for a recursive grammar. Views of the Lexicon diverge sharply for various investigators. For some it is just a list of items, where what an item can be also differ. The simplest view says that the lexicon is a list of words, but this view is problematic on several counts. We distinguish between a (morphological) word and a lexeme. A lexeme may consist of several words (including discontinuous strings or structures (*give(…) up*, *put up with*) or, possibly, parts of words (affixes).

Better to say a *listing* of items, as in some languages (e.g. perhaps English), the set of lexemes is not finite.

Some frameworks deny a principled distinction between Lexicon and Grammar, for example, perhaps Construction Grammarians (see above).

Some Questions:

What is a word? (Di Sciullo and Williams, Bach 1983)

Inflectional vs derivational morphology.

If we think of the Lexicon as a base for a recursive grammar, then we may think of derivational processes and rules as means for adding to this base, that is, forming new lexical items. We then identify inflectional morphology as word-modifying processes that are necessary for the grammar.

Features.

Most current systems use features of various kinds. They can be thought of conveniently as systems of functions from linguistic expressions to values in various domains appropriate to the class of linguistic expressions for which they are defined.
For example the Latin word *feminam* has the values [**fem, singular, accusative]** for the features **[gender, number, case]**. Spelling this out is part of the specification of the structure of k-tuples of linguistic signs

IV. Information systems

Given a grammar of some kind then an **information system** provides various procedures for manipulating, accessing, modifying etc objects (signs) of the sorts specified by the grammar. The idea of an information system is intended to be broader than just those systems dealing with language in the narrow sense.

A simple example of an information system:

Example 1: a bibliographical database

The body of the system is a list of entries. Each entry contains representations of the following sorts:

type: book, article, other
date:
authors:
editors:
publisher / journal
ISBN number
place of publication
....

Possible actions

search by key (author, title, subject, etc.)

outputs

emendations: corrections, additions, removals

Note that a bibliographical database may have or be associated with a number of different interfaces to a "world": a library, the set of all published documents, a set of desiderata for setting up a working library, or even a virtual library including planned or desired documents. And so on.

V A sample simple set up:

The basic units of the system are Information States: IS-1, IS-2,...

Each IS is accessible to an interlocutor, coded as IL_i-1, IL_i-2, etc. such that ISi is accessible to IL-i

A **common information state** is relativized to one or more interlocutors: so $CIS_{1/2}$ is the intersection of IS_1 and IS_2 (and so on).

A **dialogue** is a sequence of expressions (sentences, etc) and associated information states.

Example 2: chess

Imagine a game of chess. Two interlocutors: W and B communicate by email. In the initial state the common $CIS_{W/B}$ is a board with spaces correlated with pieces in the standard initial layout. A game is a sequence of CISW/B's beginning with the initial state and terminating with a CheckMate or Draw CIS or a CRASH (dinnertime, someone spills the board, etc.). The game is given as a series of Moves (WKP to WK-4) conforming to the syntax and semantics of the game, and the semantics is given as mapping from CS's to CS's according to the moves.

Abstraction: in the representation of a chess game, all irrelevant information is ignored. So suppose there is a representation of a real chess game played by Jones and Karnofsky in Capetown on a certain day: what is ignored: where, who watched, what time it started and ended, the termperature and airpressure (and changes of them) etc. Note the difference from a news report on that real chess game, where such details would be routinely included in a narrative.

Moving closer to the current project:

Example 3: a flight

Parameters:

aircraft (with specs)

crew: (principal pilot, copilot, crew-1, crew-2, etc)

noncrew persons (passengers, observers, etc.)

position: longitude, latitude, altitude (planet, corresponding items for intergalactic and galactic journeys)

current airspeed / groundspeed

scheduled airspeed / groundspeed

time (GMT and/or local)

other individuals: principally places

status of places (e.g. target, rendezvous point etc.)

Rules and constraints (and automatic error messages and consequences):

For example: in a flight dialogue if an information state is specified for a time earlier than the time of the current information state, then error message requests correction and refuses update of current common ground CISi,...,k.

A full information flight dialogue gives common information states for the entire crew and assumes accessibity to the entire crew. Note that some of the information to be included in the CIS's can be automatically recorded directly from the instruments of the aircraft, satellite tracking, identity of source input etc. and checked against radar info etc. (see section on logs below).

VI. Information states and systems of grammar and interpretation.

If we look at various systems mentioned above, we can ask how dialogues could be related to them.

Of all the systems mentioned, it seems that the Discourse Representation Structures (DRS) of Kamp come closest to what we intend by information states. Let us recall what DRS's are like and what they contain. Here's an example of the way DRT interprets a sequence of sentences:

1. John has a cat. He loves it.

On the interpretation in which *He* refers to John, and *it* refers to John's cat, we might have:

x y john
x = john
cat(y)
have(x,y)

z w
love(z, w)
z = x
w = y

These DRS's can be melded together to yield:

x y john
x = john
have(x, y)
cat(y)
love(x,y)

We can also adopt for convenience a more traditional kind of representation (following Chierchia 1995) with square brackets:

x, y, john[x = john & cat(y) & have(x,y)] z, w[love(z,w)] & z = x & w = y]

x, y, john[x = john & have(x,y) & cat(y) & love(x,y)]

A further reduction would substitute the constants for variables with which they are identified to give

john, y [cat(y) & have(john, y) & love(john,y)]

The embedding conditions will count this DRS as true in a situation (or world) S if it can be satisfied in S, that is, if john is in S, and there is an individual that is a cat in S and that john has and loves. In effect, then in such an example the interpretation amounts to existentially quantifying the free variables that are "left" after any other conditions are satisfied. A different result would obtain if we wanted to interpret this sentence:

If John has a cat, he loves it.

This example requires spelling out the conditions for *if ... then* sentences (see Kamp 1981, Kamp and Reyle 1993)..

Evidently, something like DRS's, suitably extended and structured, could be very directly used to create and modify Information States and further to construct interpreted dialogues and Common Information States.

The task then is twofold: (1) to show how to get from texts to information states, (2) to show how to use these information states, suitably tagged and annotated, to get to records of dialogues, etc.

More on the Structure of a DRS system:

Each DRS has the following components:

i. a set of variables and constants

ii. a set of conditions represented by formulas consisting of a predicate or relation symbol and the appropriate number of arguments, drawn from (i), where the predicate symbols include identity (x = john).

In the box representations, (i) consists of the contents of the highest subbox. In the bracket notation (i) is a sequence of comma-delimited elements in front of a left bracket.

Manipulations of DRS include amalgamation, processes by which two (or more) DRS's may be put together as in the example given above where an amalgamation obeys the identity conditions and performs appropriate substitutions in the conditions, or combined according to embedding conditions.

Let us now take a run at how DRS's can be used to implement various kinds of records and other objects built up with the Information Systems we have outlined above.

Information States and Discourse Representation Structures

An Information State can be thought of as a partial representation of a situation S. It is possible to give an IS in the form of a DRS, but we will describe the form of an IS independently.

In the first instance we want to build into an IS an explicit place for information of the sort that can be related to notions like "common ground" (Stalnaker 1978). In DRT, the set of constants in the domain that may be freely appealed to in interpretations might be thought of in this way (john in our example above). Recall that information states are thought of as correlated with ("for") particular participants in a dialogue, for example. So they are perhaps better thought of as "belief states." Elements in these belief states can be offered as candidates for general information in the common ground of the interlocutors in a dialogue.

What kinds of things are available in the (common) ground? In the first instance a set of individuals. But of course there is a whole lot of information about these individuals, which can be represented as predications. In addition, there are general constraints, physical laws, and so on. Of course, it is not practical or perhaps even possible to represent all of this as part of the specification of a system of information states. What is needed is a specification of items and information about them that is necessary for creating and understanding a particular application: narrative, log, dialogue, etc.

VII. Kinds of exchanges

Statement and Acknowledgement:

“We are at location such and such.”

“I copy you.”

Queries and Answers

“What is your current location?”

“ We are at location such and such.”

Requests and Compliances

“Proceed to location x,y.”

“OK.”

VIII. Some samples.

Sample i. A monologue.

In a monologue there is only one locutor, so we can take all the information states as being those of this single participant and dispense with indexing information states to interlocutors.

We start with a blank IS, and update after every utterance:

IS-0:

U-1 John is in London. ==> in(london)(john)(t_0)

This step is directly available from a grammar together with a procedure for enumerating all the possible analyses and interpretations that the grammar can assign to the sentence.

IS-1:

john London t_0

at t_0 [in (London, john)]

CG = IS-1

Comments: Without an explicit laying out of initial conditions, the first IS is taken to be blank. The first bit of language licenses the introduction into the domain of three entities: john, london, and a time and adds the information specified. Since there is no other interlocutor the information given in the information state IS-1 is taken as common ground (Stalnaker 1978, Lewis 1979).

Note that the information to be gleaned from this bit of narrative is incomplete. We don't know what John is intended nor what London. If this were a dialogue you might expect the interlocutor to ask for clarification: John who? or (in Canada at least) Which London? Such clarifications might of course be parts of a monologue as well: "London, Ontario, that is." "I mean John Osborne of Portage La Prairie."

Keeping track of individuals, times and locations can be made explicit for narratives. Compare here the script of a play: a list of characters, a place ("in the sitting room of house in North London"), a time ("sometime in the first half of the 20th century"), stage directions ("the following day"). In other kinds of narratives, such information is given in part by appropriate uses of the tense and aspect system of a language, in particular the use of verbs in various forms and classifications (Kamp and Reyle 1993; Kamp's work with Chr. Rohrer [reference]).

For example, suppose we replace the utterance given above by a sentence in the simple past tense:

U-1': John was in London.

With no explicit previous setting up of a time, we accommodate by introducing a time reference and interpret the sentence as referring to a state of affairs that holds at this (past) time. A stative sentence like this introduces a condition that is then assumed to hold until something in the narrative implicitly or explicitly sanctions a change.

Sentences about other kinds of eventualities, direct the interpreter to construct a sequence of events ordered by "and then." Additional stative sentences amplify the conditions on the initial state. (Kamp and Reyle 1993).

Similar considerations and options are required for keeping track of or making inferences and guesses about locations. Compare the example of a play with explicit stage directions ("in another part of the house" etc.)

The need for such strategies for reconstructing locations and times is obviated in a particular kind of text: a log.

.

Sample ii: Logs

Think of the log of a voyage. Such a log will contain explicit information about the initial conditions for the voyage: name and identifying information about the vessel, the crew, passengers, supplies, pertinent properties of the vessel. Along with this information about the initial conditions, we can imagine a detailed plan (or pointer to such a plan) for the "present" voyage.

The remainder of the log will consist of entries ordered or tagged for time and notated as to location, and other information pertinent to the plan and purpose of the voyage.

In ancient days, a voyage log would be created by hand. It required human interventions: noting times from a chronometer, taking readings for positions and so on, and entering the information into the logbook at regular intervals, Nowadays, much of this will be automated. A running record can be made for time, position, altitude, groundspeed, airspeed, and so on. Into or along with such a record, a transcript might be kept for information entered into an electronic record by hand, audio recordings and/or transcriptions of messages from the ground or among the crew, announcements and so on. From such records and recordings a sequence of Information States can be abstracted.

Now consider a log which contains records of all communications among members of a crew. Suppose there are three crewmembers. A model for a record of a flight could take the form of a time-line with channels for each crewmember supplemented with lines for time and other coordinates.

Such a log requires processing that extracts information and, in the first instance, divides the information into two types: information relevant to the goal of the voyage, and information that is not so relevant: relevant: information about targets, meeting points, sightings of objects, etc. ; irrelevant: two crew members discuss a recent date, baseball game, the US elections, etc

Efficiency will be improved by including in the system certain conventionalized utterances or bits of utterances that do not need analysis but can be directly linked to the information states: "Roger that" and the like.

Sample iii: A mission log.

Here is an outline of what might be form for keeping track of what goes on during a mission in the form of a series of Information States.

Initial conditions for the mission: time, place of origin; destination; route; identification key for mission; personnel: aircrew, groundcrew; aircraft, etc. A sequence of records of communications in the form of a database, where each record has the following fields:

TIME

LOCUTORS Originator: Receiver:

UTTERANCE transcription

SEMANTICS a translation of the utterance into some logical form (predicate argument structure supplemented with

INTerpretation: extracted and supplemented

UPDATES: TIME: LAT: LONG: ALT: AIRSPEED: and

DOM(ain): a cumulative set of all entities involved and introduced (starting from initial conditions including COMMON GROUND) and adding entities introduced in dialogue or other sources of information

ETC

The flavor of such a series of records can be gleaned from the first few entries in a example drawn from Beth L.'s version of one of the mission records:

The INT fields are just lifted directly from Beth L.'s version.

REF: mission1

TIME: 00:05.0

LOCUTOR: EXP

UTTERANCE: Okay team, there's the start of your first mission. Good luck.

SEM: [Okay team] start(your (first (mission)) [wish (good (luck))]; you= ??

INT: Begin

UPDATE: TIME: 00:05.0 LONG: LAT: ALT: AIRSPEED: DOM: (TEAM EXP)

ETC:

ID: mission1:EXP:00:05.0

TIME: 00:24.0

LOCUTOR: AVO

UTTERANCE: Hey, when you have your first point, just let me know.

SEM: [Hey: attention!] when(have(first(point))(YOU)!let(know ((have(first(point)))(YOU)) (AVO))

INT: Ready to receive Point 1

UPDATE: TIME: 00:24.0 LONG: LAT: ALT: AIRSPEED: DOM: (AVO TEAM EXP)

ETC:

ID: mission1:AVO:00:24.0

TIME: 00:28

LOCUTOR: DEMPC

UTTERANCE: Okay, AVO this is Dempc, our first waypoint will be LVN. It's the roz entry point.

SEM: [Okay] to AVO: speaking DEMPC: WILL[OUR (first (waypoint))(LVN) & BE (the (roz (entry_point)))(LVN)

INT: Point 1 is LVN. Roz entry point.

UPDATE: TIME: 00:28 LONG: LAT: ALT: AIRSPEED: DOM: (DEMPC AVO TEAM EXP)

ETC:

ID: mission1:DEMPC:00:28

TIME: 01:08.0

LOCUTOR: AVO

UTTERANCE: Are there any restrictions there?

SEM: ?THERE(there)(restrictions); there = LVN

INT: Query Restrictions

UPDATE: TIME: 01:08.0 LONG: LAT: ALT: AIRSPEED: DOM: (LVN DEMPC AVO TEAM EXP)

ETC:

ID: mission1:AVO:01:08.0

IX. Some Options

In any kind of information exchanging IT system there are two strategies: one is to shoot for a very general system that could handle a wide variety of types of queries without any "precooking" or restrictions, the other is to have strictly limited kinds of queries and answers. In effect, the users have to learn a relatively small sublanguage of the matrix language and stick to it for successful passing of information, answering of queries and so on. In the task at hand, probably both these kinds of approaches should be tried out and evaluated. With a view to doing this task, it seems reasonable to move into a data collection phase by undertaking a cataloguing of a large body of records to see what actually has to be handled or simulated in a training regimen.

X. Where to go from here

In view of the last considerations, then, an immediate task is to go through as many mission records as possible and assemble a description which catalogues:

Vocabulary: typical names for entities, locators, locations, etc.

Attributes and information: e.g. restrictions on airspeed, altitude, classification into targets, waypoints, and the like.

A small grammar laying out the sentence patterns, question answer pairings, responses (“copy” “Roger that”) and so on that are actually found in the records.

Testing out the results by manufacturing utterances and doing some experiments on success of information transmission, reliability of responses and so on.

The first four of these activities are basically just what a field linguist does when he is undertaking the description of a language. The nature of the task dictates a “no cheating” approach. Success has to be measured in the extent of coverage of the actual material surveyed.

Implementation questions remain and will have to be dealt with when planning actual uses of the system with various input and output options,.

References:

Ajdukiewicz, Kasimierz. 1935. Die syntaktische Konnexität. Studia Philosophica 1: 1-27. Translated as `Syntactic connexion' in Storrs McCall, ed., Polish Logic: 1920-1939 (Oxford: Oxford University Press, 1967).

Bach, Emmon. 2005. Eventualities, grammar, and linguistic diversity. In Henk J. Verkuyl, Henriette de Swart, and Angeliek van Hout, eds., *Perspectives on Aspect* (Dordrecht: Springer) pp. 167-180.

Bach, Emmon.1981.On time, tense, and aspect: an essay in English metaphysics. In Peter Cole (ed.), *Radical Pragmatics,* New York: Academic Press, 63-8l.

Bach, Emmon. 1986. Natural language metaphysics. In R. Barcan Marcus, G.J.W. Dorn and P. Weingartner, eds. *Logic, Methodology, and Philosophy of Science* VII (Amsterdam et al.: North Holland, 1986), pp. 573-595. LING

Bach, Emmon. 1994. The semantics of syntactic categories: a cross-linguistic perspective. In John Macnamara and Gonzalo E. Reyes, eds., *The Logical Foundations of Linguistic Theory* (New York and Oxford: Oxford University Press), pp. 264--281.

Bach, Emmon. 1980. Tenses and aspects as functions on verb-phrases. in C. Rohrer (ed.), *Time, Tense and Quantifiers: Proceedings of the Stuttgart Conference on the Logic of Tense and Quantification*. Pp. 19-38.

Bach, Emmon. 1986. The algebra of events. *Linguistics and Philosophy* 9 (1986): 5-16. Reprinted in Paul Portner and Barbara H. Partee, eds. 2002. Formal Semantics: the Essential Readings. [Oxford] Blackwell. Reprinted in Inderjeet Mani, James Pustejovsky, and Rob Gaizauskas, eds., *The Language of Time: a Reader*. (2005: Oxford: Oxford University Press), pp. 61-69.

Bach, Emmon and Wynn Chao. Semantic universals and typology. In press.To appear in Chris Collins, Morten Christiansen and Shimon Edelman, eds., *Language Universals* (Oxford: Oxford University Press).

Chierchia, Gennaro. 1995. *Dynamics of Meaning*. Chicago: University of Chicago Press.

Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, Mass. and London: The MIT Press.

Cinque, Guglielmo, and Richard S. Kayne, eds. 2005. *The Oxford Handbook of Comparative Syntax*. Oxford: Oxford University Press.

Culicover, Peter W., and Ray Jackendoff. 2005. *Simpler Syntax*. Oxford University Press.

Cresswell, Max J. 1973. *Logics and Languages*. London: Methuen.

Deacon, Terrence W. 2003. Universal grammar and semiotic constraints. In Morten H. Christiansen and Simon Kirby, eds. *Language Evolution*. Oxford: Oxford University Press. pp. 111-139..

Dowty, David R. 1978.Lexically governed transformations as lexical rules in a Montague Grammar. *Linguistic Inquiry* 1: 45-78.

Dowty, David R. 1979. *Word Meaning and Montague Grammar*.
Dordrecht: Reidel.

Dowty, David. 2007. Compositionality as an Empirical Problem. In Chris Barker and Pauline Jacobson, eds., *Direct Compositionality* (Oxford: Oxford University Press), pp. 23-101.

Fillmore, Charles, Paul Kay and Catherine O'Connor. 1988. Regularity and Idiomaticity in Grammatical Constructions: The Case of let alone. *Language* 64: 501-38.

Goldberg, Adele E. 1995. Constructions: a Construction Grammar Approach to Argument Structure. Chicago and London: The University of Chicago Press.

Groenendijk, Jeroen, and Martin Stokhof. 1991. Dynamic predicate Logic. *Linguistics and Philosophy* 14:39-100.

Groenendijk, Jeroen, and Martin Stokhof. 1990. Dynamic Montague Grammar. In L. Kalman and L. Polos, eds. *Papers from the Second Symposium on Logic and Language. Budapest*: Akademiai Kaido;.

Heim, Irene R. 1982. The Semantics of Definite and Indefinite Noun Phrases. Ph.D. dissertation: University of Massachusetts, Amherst.

Heim, Irene. 1983. File change semantics and the familiarity theory of definiteness. In Rainer Bäuerle, Christoph Schwarze and Arnim von Stechow, eds., *Meaning, Use, and the Interpretation of Language* (Walter de Gruyter), pp. 164-190. [Reprinted in Portner and Partee 2002: pp. 223-248.]

Heim, Irene and Angelika Kratzer. 1998. *Semantics in Generative Grammar*. Oxford: Blackwell.

Hockett, Charles F. 1963, 1966.The problem of universals in language. In Joseph H. Greeberg, ed. *Universals of Language: Second Edition* (Cambridge, Mass: MIT Press), pp. 1-29.

Jackendoff, Ray. 1990. *Semantic Structures*. Cambridge, Massachusetts:MIT Press

Kamp, H. 1981. A Theory of Truth and Semantic Representation. In J. Groenendijk, T. Janssen, and M. Stokhof, eds., *Formal Methods in the Study of Language*, Part 1: 277-322 (Amsterdam: Mathematical Centre Tracts 135).[Reprinted in Portner and Partee 2002: pp. 189-222.]

Kamp, H. and U. Reyle. 1993. *From Discourse to Logic*.Dordrecht: Kluwer.

Kratzer, Angelika. 1989. An investigation of the lumps of thought. *Linguistics and Philosophy* 12: 607-653.

Kratzer, A. 2007. Situations in Natural Language Semantics", The Stanford Encyclopedia of Philosophy (Spring 2007 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2007/entries/situations-semantics/>

Alain LeComte. 2005. Categorial Grammar for Minimalism, Language and Grammar: Studies in Mathematical Linguistics and Natural Language, Casadio, Scott, Seely eds. CSLI Lecture Notes number 168, pp 163-188, PDF, Stanford, USA, 2005

Lecomte, Alain, and Christian Retoré 2001. Extending Lambek grammars: a logical account of minimalist grammars. ACL 2001: 354-361.

Lewis, David. 1979. Scorekeeping in a Language Game. Rainer Bäuerle, Urs Egli, and Arnim von Stechow, eds., *Semantics from Different Points of View*. Dordrecht: Springer-Verlag. Reprinted in Portner and Partee 2002.

Montague, Richard. 1970.Universal Grammar. [Paper 7 in Montague, 1974.]

Montague, Richard. 1973. The Proper Treatment of Quantification in Ordinary English. Paper 8 in Montague, 1974, originally published in Hintikka, Moravcsik, and Suppes, 1973, p. 221-42.]

Montague, Richard. 1974. *Formal Philosophy*. Edited by Richmond H. Thomason. New Haven: Yale University Press

Oehrle, Richard T., Emmon Bach, and Deirdre Wheeler, eds. 1988. Categorial Grammars and Natural Language Structures. Dordrecht / Boston / Lancaster / Tokyo: Reidel.

Stabler, Edward P. 1997. Computing quantifier scope. In Anna Szabolcsi, ed., *Ways of Scope Taking* (Dordrecht / Boston / London: 1997), pp. 155-182.

Stalnaker, R. 1978. Assertion. *Syntax and Semantics* 9, 315-332.

Steedman, Mark. 1996. *Surface Structure and Interpretation*. Cambridge, Massachusetts: MIT Press.

Steedman, Mark. 2000. *The Syntactic Process*. Cambridge, Massachusetts / London: MIT Press.